JOURNAL OF VIROLOGY, May 1997, p. 3666–3676 Vol. 71, No. 5
0022-538X/97/$04.00+0

Complete Nucleotide Sequence of Simian Endogenous Type D Retrovirus with Intact Genome Organization: Evidence for Ancestry to Simian Retrovirus and Baboon Endogenous Virus

ANTOINETTE C. VAN DER KUYL,* RUI MANG, JOHN T. DEKKER, AND JAAP GOUDSMIT
Department of Human Retrovirology, Academic Medical Centre, 1105 AZ Amsterdam, The Netherlands

Received 18 November 1996/Accepted 5 February 1997

A complete endogenous type D viral genome has been isolated from a baboon genomic library. The provirus, simian endogenous retrovirus (SERV), is 8,393 nucleotides long and contains two long terminal repeats and complete genes for *gag*, *pro*, *pol*, and *env*. The primer binding site is complementary to $tRNA_3^{Lys}$, like in lentiviruses. The *env* GP70 protein is highly homologous to that of baboon endogenous virus (BaEV). PCR analysis of primate DNA showed that related proviral sequences are present in Old World monkeys of the subfamily *Cercopithecinae* but not in apes and humans. Analysis of virus and host sequences indicated that the proviral genomes were inherited from a common ancestor. Comparison of the evolution of BaEV, exogenous simian retrovirus types 1 to 3 (SRV1 to SRV3), and SERV suggests that SERV is ancestral to both BaEV and the SRVs.

Retrovirus sequences have been found in the genomic DNAs of many species of mammals, including primates. Most of these sequences represent ancient integrations, rarely complete viral genomes, and even if that is the case, the genomes often can no longer be expressed due to the presence of multiple stop codons and deletions. Complete endogenous viruses described in primates include type C viruses from macaques, the owl monkey, and *Colobus polykomos* (34, 36, 42, 44), the chimeric type C/type D baboon endogenous virus (BaEV) in baboons and other species of African monkeys (3, 46), and possibly human endogenous retrovirus K10 in humans (6), which is most closely related to type B retroviruses. Endogenous type D sequences have been observed in langurs (*Presbytis* species) and squirrel monkeys (4, 16). No complete sequences have been determined for these type D viruses, and it has also not been proven that the infectious langur isolate Po-1-Lu and the exogenous squirrel monkey retrovirus (SMRV) are actually derived from endogenous type D sequences. Exogenous type D viruses (simian retroviruses, or SRVs) have been found in captive macaques worldwide. Several neutralization serotypes are recognized (13), of which SRV-1, SRV-2, and SRV-3 (Mason-Pfizer monkey virus) are the best characterized, and complete nucleotide sequences have been determined (33, 39, 41). These viruses induce a fatal immunodeficiency syndrome in rhesus macaques (initially called SAIDS for simian AIDS [28, 40]), but SRV is unrelated to the simian immunodeficiency viruses, which are currently recognized as the simian counterparts of the human immunodeficiency viruses. Recently, an exogenous type D virus, SRVpc, in apparently healthy captive baboons has been described (14). A high percentage of West African wild-caught talapoin monkeys were found to have antibodies to type D viruses, suggesting that talapoins could be a natural reservoir of SRVs (18). Humans do not appear to be hosts for any type D viruses, except maybe in extremely rare instances of incidental zoonotic infections. Extensive serological surveys have shown no proof of infection (13), even in people who have had close contact with infected primates. To date, only a single human isolate of a type D virus, from a severely immunosuppressed patient with AIDS, has been reported (5).

Recently, we have studied the structure of BaEV genomes integrated into the genome of the yellow baboon (*Papio cynocephalus*) by hybridizing a genomic library to BaEV-specific probes. Most lambda phages obtained contained sequences identical to that of BaEV, but two phages, designated 25.2 and 23.1, which hybridized only weakly to a BaEV *env* probe and not at all to BaEV reverse transcriptase (RT) and long terminal repeat (LTR) probes, were found (45). These two lambda clones could be PCR amplified with BaEV *env* primers, and the obtained fragments showed approximately 80% nucleotide (nt) homology to BaEV. Closer inspection of the lambda clones revealed the presence of *gag*-like sequences with approximately 90% homology to SRV.

In this study we have analyzed the genome organization, copy numbers, and host distribution of this new intact endogenous type D retrovirus which is the putative ancestor of both BaEV and the pathogenic SRVs.

MATERIALS AND METHODS

Baboon genomic library. A baboon genomic library in the lambda DASH^{R}II vector was obtained from Stratagene (La Jolla, Calif.). This library was constructed from kidney tissue of a normal 18-year-old male *Papio cynocephalus* baboon.

Isolation of type D proviral clones. The baboon genomic library was screened by using a ^{32}P-labelled probe homologous to the BaEV *env* sequence (45).

Lambda DNA was isolated from purified positive plaques by using the Wizard Lambda Preps DNA purification system from Promega (Madison, Wis.).

Primate samples. Samples were obtained from the following African primates: *Macaca sylvanus* (barbary monkey), *Mandrillus sphinx* (mandrill), *Papio cynocephalus cynocephalus* (yellow baboon), *Papio cynocephalus anubis* (olive baboon), *Papio cynocephalus ursinus* (chacma baboon), *Papio hamadryas hamadryas* (sacred baboon), *Theropithecus gelada* (gelada), *Cercocebus torquatus atys* (sooty mangabey), *Cercocebus aterrimus* (black mangabey), *Cercocebus galeritus* (tana mangabey), *Cercopithecus aethiops aethiops* (grivet), *Cercopithecus aethiops pygerythrus* (vervet), *Cercopithecus aethiops tantalus* (tantalus monkey), *Cerco-*

* Corresponding author. Mailing address: Department of Human Retrovirology, Academic Medical Centre, Meibergdreef 15, 1105 AZ Amsterdam, The Netherlands. Phone: 31 20 566 4522. Fax: 31 20 691 6531. E-mail: Kuyl@amc.uva.nl.

pithecus aethiops sabaeus (green monkey), *Cercopithecus diana roloway* (diana monkey), *Cercopithecus mitis* (blue monkey), *Cercopithecus mona pogonias* (mona monkey), *Cercopithecus patas* (patas monkey), *Cercopithecus ascanius* (red-tailed monkey), *Cercopithecus nictitans* (spot-nosed guenon), *Cercopithecus cephus* (moustached monkey), *Cercopithecus neglectus* (De Brazza's guenon), *Cercopithecus hamlyni* (owl-faced guenon), *Pan troglodytes* (common chimpanzee), *Pan paniscus* (pygmy chimpanzee), *Gorilla gorilla* (gorilla), and *Homo sapiens* (human). Additional samples were obtained for three Asian monkey species: *Macaca mulatta* (rhesus macaque), *Macaca nemestrina* (pig-tailed macaque), and *Macaca fascicularis* (crab-eating macaque).

The origins of the samples (serum, plasma, and/or blood cells) were as published before (46). *Cercopithecus hamlyni* DNA was obtained from John Hancock (MRC Clinical Sciences Centre, London, United Kingdom). Additional samples of *Papio hamadryas* genomic DNA were obtained from the Zoologischer Garten Leipzig (Leipzig, Germany) through the European Gene Bank of Primates (Munich, Germany), which also supplied their DNA concentrations.

DNA extraction and PCR amplification. Total DNA was extracted by a procedure using silica and GuSCN (7). A 764-nt fragment from this DNA was amplified with a primer set consisting of an upstream primer, 5′ GGCATTCC CTACAATCCCC 3′, located in the *pol* gene and a downstream primer, 5′ GGCAATAAAACATACTATACG 3′, located in the *env* gene. PCR amplifications were performed by the following protocol: denaturation for 5 min at 95°C, amplification for 35 cycles of 1 min at 95°C, 1 min at 55°C, and 2 min at 72°C, and then an extension of 10 min at 72°C. Obtained fragments were cloned into the TA vector (Invitrogen, San Diego, Calif.) and sequenced. At least three clones from each individual were analyzed.

DNA sequencing. Sequencing of the clones with an Applied Biosystems 373A automated sequencer was done in both directions directly from lambda DNA with purified specific primers, following the manufacturer's protocols. To resolve a few ambiguities, some sequencing primers were also used to generate PCR fragments from the lambda clones for additional sequencing. PCR fragments obtained from monkey genomic DNA were sequenced from plasmid DNA with T7 and M13 primers.

Sequence analysis. Alignment of the sequences was done with CLUSTAL (17). The phylogenetic analyses were done by the neighbor-joining (NJ) method, as implemented in the MEGA package (23). Distances were estimated by Kimura's two-parameter method (21) for nucleotide sequences and by the p-distance method for protein sequences. One hundred bootstrap replicates were analyzed. Using other methods for the determination of nucleotide sequence distances did not influence the trees. Gaps introduced for optimal alignment were not considered informative and were not included in the analyses. The GenBank accession numbers of the comparison sequences were M11841 (SRV1), M16605 (SRV2), M12349 (SRV3), D10032 (BaEV), M26927 (gibbon ape leukemia virus [GALV]), M23385 (SMRV-H), U16843 (SRVpc P27), and U16844 (SRVpc GP20).

Copy number determination. Copy numbers of simian endogenous retrovirus (SERV) proviruses in the baboon genome were estimated by limited dilution and nested PCR (31). The outer primers for the PCR are described above, and the nested primers used were 5′ GATGCAGCCAGATGGCTGCC 3′ (upstream) and 5′GCTGTCTTGCCCGAGCAAGTC 3′ (downstream), which together amplify a fragment consisting of 272 nt. The nested PCR was optimized to amplify a single copy of input DNA. Baboon genomic DNA of known concentration was diluted in 10-fold steps for the first PCR, and the last positive sample was used for additional 2-fold dilution steps. For each 2-fold dilution step, 10 nested PCR reactions were performed twice.

Nucleotide sequence accession numbers. The sequences reported in this paper have been deposited in the GenBank data base (accession no. U85505 and U85506).

RESULTS

Nucleotide sequences of the proviruses. The baboon genomic clone 23.1 contains a complete retrovirus sequence of 8,393 nt with a genomic organization identical to that of known type D viruses. The base composition of the genomic viral RNA can be deduced to be 31.4% A, 26.0% U, 23.9% C, and 18.7% G. Four main open reading frames (ORFs) are present, with coding regions for *gag*, *pro*, *pol*, and *env* proteins. The protease gene is overlapping both the *gag* and *pol* genes, in the same way as reported for type D viruses. The coding region is flanked by LTRs of 484 and 479 nt.

Baboon genomic clone 25.2 contains a related proviral sequence which is missing the 5′ LTR and the first part of the *gag* gene due to the method used for construction of the library. The remaining 7,113 nt of viral sequence has a high degree of homology to the clone 23.1 virus, including a similar 3′ LTR.

The LTR and regulatory elements. Viral sequence 23.1 contains two LTRs; the 5′ LTR is 484 nt long, while the 3′ LTR is only 479 nt in length (Fig. 1). This difference can be attributed to a deletion of ATAAT in an A-T stretch in the 3′ LTR. The 3′ LTR of virus 25.2 is also 484 nt long, suggesting that this size is the original length. SRV LTRs sequenced to date are 345 to 346 nt in length.

The LTR boundaries are defined by the inverted repeat TGTCC/GGACA, like in other type D viruses, but the left repeat has been mutated to GGATA in both LTRs of clone 23.1. Although the LTR sequences are difficult to align with SRV LTR sequences, except for the 3′-most 140 nt, a GenBank search found significant homology only with type D LTRs, while no similarity was observed with other retroviruses or transposons.

The 5′ and 3′ LTRs of virus 23.1 are not identical but have a total difference of 18 nt, in addition to the earlier-mentioned deletion, suggesting that this provirus integrated in the monkey genome a long time ago. More differences are observed between the LTRs of clone 23.1 and the single LTR of clone 25.2. The LTRs of the 23.1 provirus are flanked by identical GTG GCA genomic repeats.

The 5′ LTR of 23.1 is followed by a primer binding site (PBS) complementary to the 3′ end of $tRNA_3^{Lys}$, which is unusual in type D viruses as they commonly utilize $tRNA_{1,2}^{Lys}$ to prime their minus-strand synthesis. $tRNA_3^{Lys}$ is used by the lentivirus group (e.g., visna virus, HIV, SIV) and by mouse mammary tumor virus, a type B virus. A polypurine tract (PPT), involved in plus-strand synthesis, precedes the 3′ LTR of each of the two viruses. The two PPTs are identical except for the deletion of a single A residue in the clone 23.1 PPT.

Another regulatory element present in the 5′ untranslated region is the packaging or encapsidation signal Ψ. For SRV3, biochemical studies have indicated that the sequence between the PBS and the splice donor site before the start codon of *gag* is important and shows extensive secondary structure (15). Alignment of SRV packaging elements with clone 23.1 showed that although the overall homology is only 46%, a stem-loop structure predicted in 10 additional retrovirus 5′ leader regions is completely conserved in the 23.1 sequence, including the AAC triplet at the top of the loop (Fig. 1B). Also, the splice donor site is present in clone 23.1.

A constitutive transport element (CTE), involved in nuclear transport and splicing of intron-containing mRNA, is present between the *env* gene and the PPT of SRV3 (8). The element consists of an imperfect repeat which can be folded into a hairpin loop. There are strongly conserved domains present, as evidenced by comparing the SRV CTE sequences with those of 23.1 and 25.2 (Fig. 1C), suggesting that homologous structures exist in these viruses, although a large deletion is observed in the 23.1 CTE. A more recent study suggested that sequences extending into the 3′ LTR are also involved in CTE function and form a third hairpin structure (35), but this observation is not supported by analyses of the homologous sequences from other SRV isolates, 23.1, or 25.2.

Viral proteins. (i) *gag*. The *gag* protein of clone 23.1 is probably translated from the first available ATG codon at position 640, although an ORF is present in front of this start codon. The *gag* precursor protein is 658 amino acids long, which is comparable to the length of other type D *gag* precursors, and the translation product can be easily aligned with *gag* proteins from SRV1, SRV2, and SRV3 (Fig. 2A), but there are two regions of extensive amino acid mismatches.

In clone 25.2, the first 222 codons of *gag* are missing but the remainder of the protein is homologous, although not identical, to the 23.1 product. The *gag* gene is completely open in

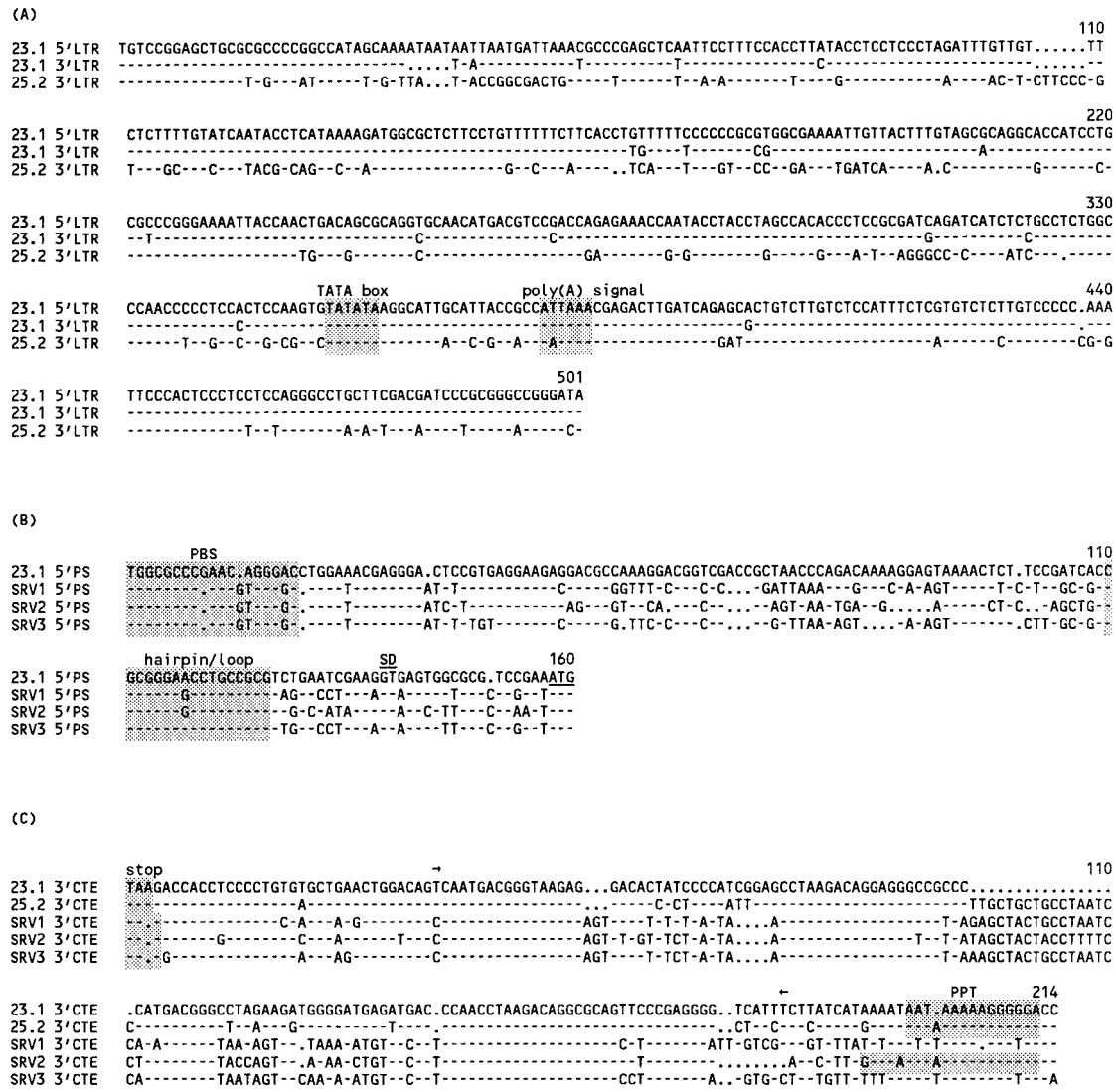

FIG. 1. (A) Alignment of the 5′ and 3′ LTRs from type D virus isolates 23.1 and 25.2. The LTRs of virus 23.1 differ in length due to a ATAAT deletion in the 3′ LTR. The TATA box and poly(A) signal are indicated. (B) Alignment of the 5′ untranslated leader sequences (between the PBS and the start codon of *gag*) of clone 23.1 (nt 487 to 642) and SRV1 to SRV3. The PBS and a conserved hairpin-loop structure have been shaded. The splice donor site (GT) is indicated (SD), while the *gag* start codon has been underlined. (C) Alignment of the 3′ regulatory CTE sequences of the 23.1 and 25.2 isolates with the SRV CTEs. The stop codon of the *env* gene is indicated, as is the putative PPT when easily identified. The arrows indicate the borders of an imperfect repeat which can be folded into a single stem-loop structure and may represent the main functional element. The sequence shown corresponds to nt 7725 to 7914 of the 23.1 sequence (between the *env* stop codon and the 3′ LTR). Gaps introduced for optimal alignment are indicated by dots; identical nucleotides are indicated by dashes.

clone 23.1, but in clone 25.2 a premature termination codon is present due to a G-to-A mutation at position 464. No large differences between the two clones are present in the downstream part of the *gag* ORF, and the 23.1 *gag* gene terminates at a TAA codon present at positions 2614 to 2616, analogous to SRV *gag* genes.

The 23.1 *gag* sequence has been analyzed by the NJ method together with a set of primate type C and type D retroviruses (Fig. 2B). GALV is included as a true primate type C virus. From this tree it is clear that the 23.1 *gag* gene clusters with those of SRV1, SRV2, and SRV3 and is more distantly related to that of the New World virus SMRV, but all are type D retroviruses. Because the complete sequence of SMRV from monkeys is not known, an SMRV sequence obtained from a human cell line (SMRV-H) was used in the analyses (30). SMRV-H is supposed to be 99% homologous to the monkey isolate and is most probably a laboratory contaminant. Figure 2C shows that the partial *gag* sequence of the baboon isolate SRVpc is very closely related to the 23.1 *gag*.

(ii) *pro*. A putative protease gene encoding a protein of 314 amino acids overlaps both the *gag* and *pol* genes in a fashion similar to that described for SRV. The gene (nt 2433 to 3377 in clone 23.1) is completely open for translation, and those of the two clones are highly homologous (95% at the amino acid level). A large degree of homology (approximately 80%) with the protease genes of SRV1, -2, and -3, which are exceptionally long among retroviruses, can be observed (Fig. 3A). The protease gene is probably expressed by translational frameshifting leading to a *gag-pro* precursor protein as proposed for SRV and Rous sarcoma virus (19, 39).

Type D viruses are known to express dUTPase activity (11). By comparison with other dUTPase genes, it has been determined that the enzyme is probably encoded by the 5′ end of the protease gene (29). Most likely, the 3′ part of the gene repre-

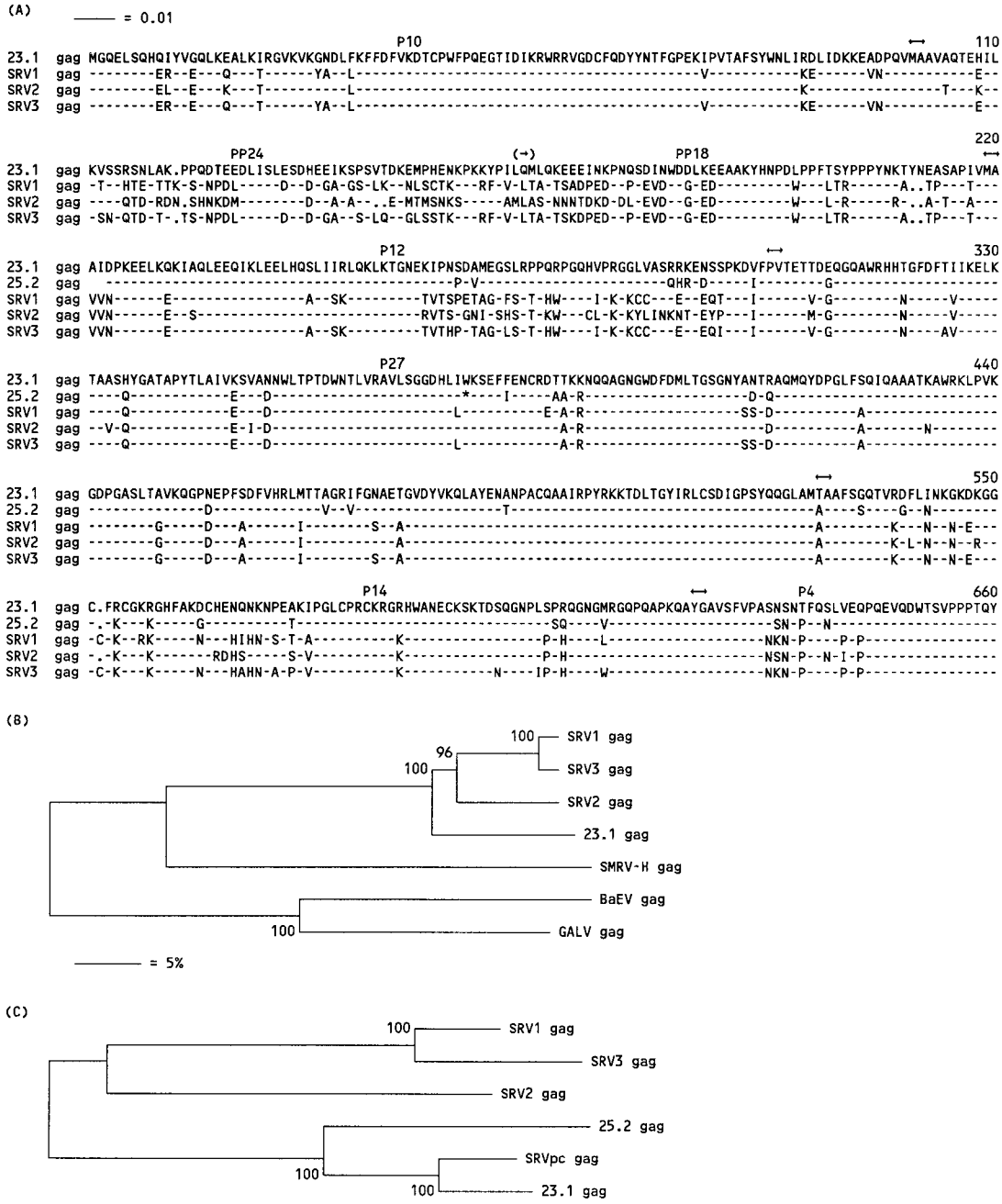

FIG. 2. (A) Alignment of *gag* proteins derived from type D virus isolates 23.1 and 25.2 and exogenous type D virus isolates SRV-1, SRV-2, and SRV-3 (Mason-Pfizer monkey virus). The first 222 amino acids are missing from the *gag* sequence of clone 25.2. The premature stop codon present in the 25.2 sequence is indicated by an asterisk. Derived major proteins are indicated by arrows (39); PP24 and PP18 are differently processed phosphoproteins, and P27 is the major core protein. The sequence shown corresponds to nt 640 to 2616 of the 23.1 sequence. (B) NJ tree based upon derived protein sequences for the *gag* genes of clones 23.1 and 25.2 and primate type C and type D viruses. Bootstrap values for 100 replicated trees are indicated. SRV1 to SRV3 are exogenous Old World type D viruses, SMRV-H is a New World type D virus, GALV is an exogenous type C virus, and BaEV is a chimeric type C/type D endogenous virus. (C) NJ tree based upon a 417-nt *gag* fragment showing the phylogenetic relationship among Old World type D viruses, including the baboon isolate SRVpc (14). Gaps introduced for optimal alignment are indicated by dots; identical amino acids are indicated by dashes.

sents the actual aspartyl protease, which would also explain the unusual length of the protease genes in type D viruses. In many lentiviruses, a dUTPase is encoded by the *pol* gene (11), and for feline immunodeficiency virus it has been shown that disruption of dUTPase expression leads to an increased viral mutation frequency (25). NJ analysis of type D protease genes shows that clones 23.1 and 25.2 are closely related to SRV1 to SRV3 and that SMRV is more distantly related to all Old World type D viruses (Fig. 3B).

(iii) *pol.* The third large reading frame present in both clones is the putative *pol* gene (nt 3356 to 5983 in clone 23.1). An alignment of the derived proteins with SRV *pol* proteins is

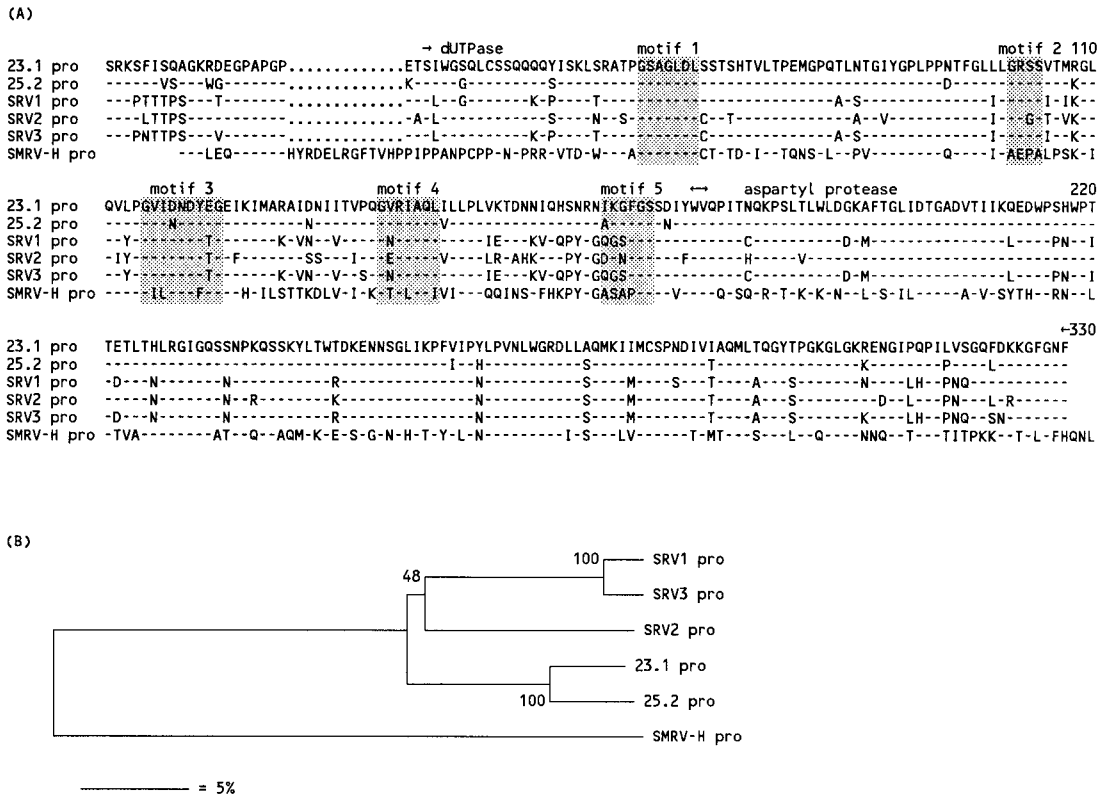

FIG. 3. (A) Alignment of protease proteins derived from type D virus isolates 23.1 and 25.2 and exogenous type D viruses SRV-1, SRV-2, and SRV-3. Codons 23 to 153 encode a putative dUTPase, while the remaining C-terminal protein sequence of 161 amino acids probably represents the actual protease. The protease gene is overlapping both the *gag* and *pol* genes in the viral genome. Close inspection of the SMRV-H sequence revealed a putative dUTPase gene upstream of its protease sequence, which is also shown. The five motifs supposedly conserved among putative dUTPase sequences (29) have been shaded. (B) NJ tree based upon type D dUTPase-protease amino acid sequences. Bootstrap values for 100 replicated trees are indicated. Gaps introduced for optimal alignment are indicated by dots; identical amino acids are indicated by dashes.

shown in Fig. 4A. Although no large deletions or insertions are present, the ORF is interrupted by two frameshifts (insertion of two single A residues at nt 5051 and 5145, respectively) in clone 23.1 and by a premature termination codon at position 4356 (G-to-A mutation) combined with a frameshift at the 3′-terminal end due to the deletion of an A residue in an A-rich stretch (nt 4580 to 4586) in the 25.2 sequence. Repeated sequencing of these positions, together with sequencing of specific PCR fragments derived from the original lambda clones, confirmed that the observed insertions and deletions of A nucleotides are actually present in the genomic clones. Investigation of homologous fragments which were PCR amplified from the DNA of baboons and other monkeys showed that the substitution leading to the premature stop codon and the A deletion leading to the frameshift were no longer observed, suggesting that in related integrations the *pol* gene is uninterrupted. Translation of both *pol* genes, ignoring frameshifts and the premature stop codon, showed that the reading frame is completely intact and homologous to those of SRV *pol* proteins. The only length differences are in the beginning, as 23.1 and 25.2 *pol* genes start with the second residue (compared to SRV), and at the 3′ end, where both reading frames continue for an additional 4 amino acids.

An NJ tree generated by using derived amino acid sequences for the *pol* genes is shown in Fig. 4B. Two distinct clusters can be observed, one containing the type C *pol* genes of GALV and BaEV and the second containing all type D *pol* genes, in which the Old World type D viruses, including clones 23.1 and 25.2, are more closely related to each other than to the New World type D virus.

(iv) *env*. Proviral clones 23.1 and 25.2 were obtained by hybridizing a genomic library with a BaEV *env* (GP70) probe. As expected, the 23.1 and 25.2 *env* genes are very similar to BaEV *env* (approximately 80% nt homology), except for the 3′-most 32 amino acids, which are more related to the type D sequence and are almost identical to each other (90% homology at the nt level). So far, no viruses have been found to have GP70 proteins with any homology to BaEV GP70. There is, however, a high level of similarity in the P20 transmembrane proteins encoded at the 3′ ends of the *env* genes of SRV and BaEV. An alignment of *env* genes is shown in Fig. 5A. Although the 23.1 and 25.2 sequences share no obvious homology with type D viruses in the GP70 protein coding region, alignment is easy in the 3′ part of the gene, which contains the highly conserved putative immunosuppressive peptide described for SRV (39) and the avian reticuloendotheliosis-associated virus. Both reading frames are interrupted by premature stop codons; clone 23.1 *env* truncates after 221 amino acids (mutation of C to A at nt 6680), and 25.2 *env* contains a nonsense mutation at nt 5967 (mutation of C to A), leading to a protein of 418 amino acids (BaEV *env* contains 563 amino acids). Phylogenetic analysis of derived amino acid *env* sequences showed a different pattern than the trees of Fig. 2 to 4. Although the BaEV *gag* and *pol* genes cluster with those of the type C virus GALV, the *env* gene clusters with type D *env* genes, confirming the hybrid nature of BaEV. The BaEV *env*

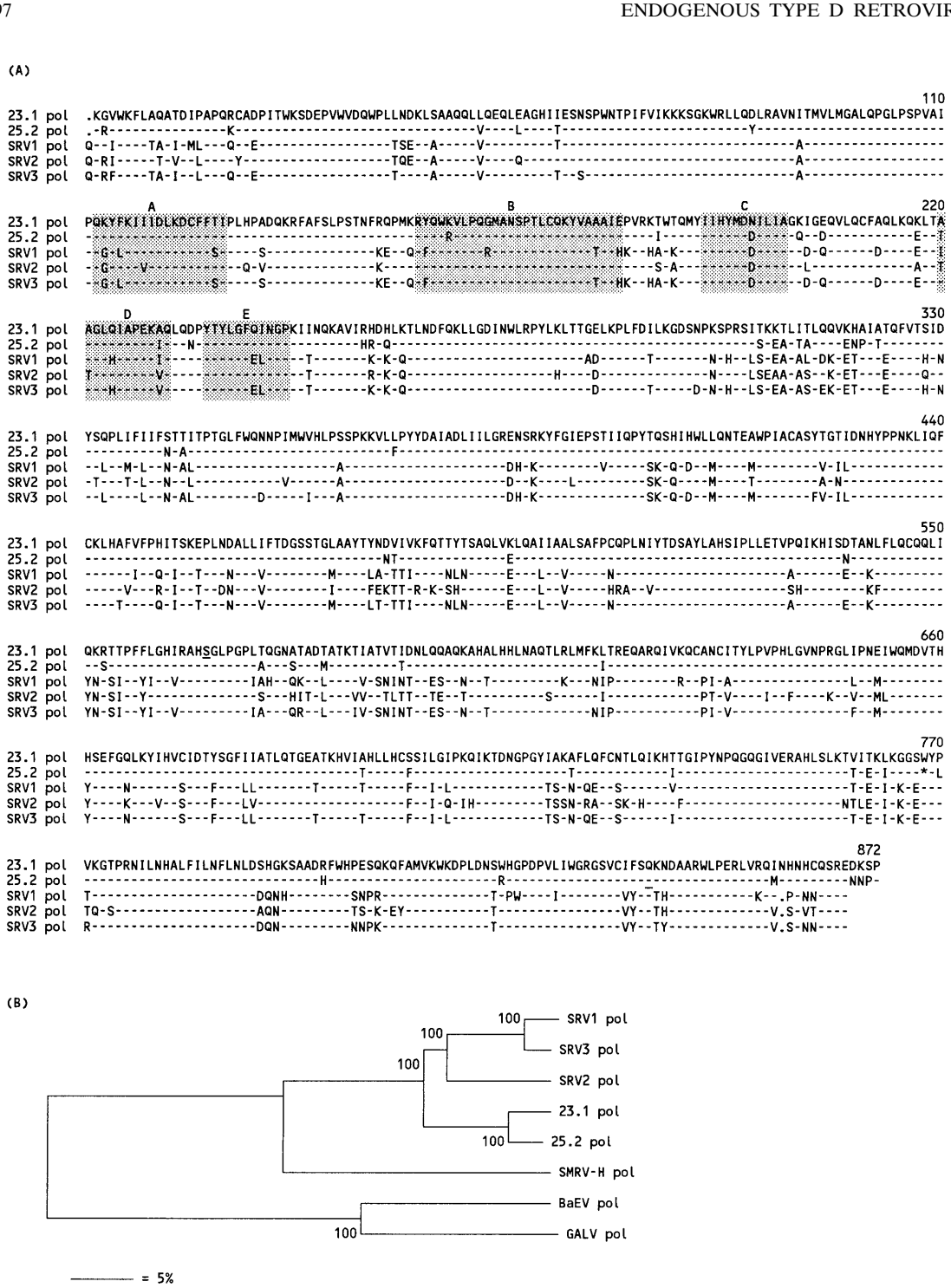

FIG. 4. (A) Alignment of *pol* proteins derived from type D virus isolates 23.1 and 25.2 and the exogenous type D virus *pol* proteins of SRV-1, SRV-2, and SRV-3. Conserved RT motifs A to E (32) have been shaded. The D-to-N mutation in motif C of clone 23.1 will probably render the RT inactive. The *pol* gene encodes both the RT and the integrase, but the precise boundaries are not known. The sequences shown for the endogenous clones have been translated ignoring the insertion and deletion of A nucleotides leading to frameshifts. The positions of the frameshifts are underlined. A premature stop codon present in clone 25.2 is indicated by an asterisk. (B) NJ tree based upon derived protein sequences for the *pol* genes of clones 23.1 and 25.2 and primate type C and type D viruses. Bootstrap values for 100 replicated trees are indicated. SRV1 to 1-3 are exogenous Old World type D viruses, SMRV-H is a New World type D virus, GALV is an exogenous type C virus, and BaEV is a chimeric type C/type D endogenous virus. Gaps introduced for optimal alignment are indicated by dots; identical amino acids are indicated by dashes.

gene is most closely related to clone 23.1 and clone 25.2 *env* genes, as could already be seen from Fig. 5A. The exogenous baboon type D isolate SRVpc is closely related to clone 23.1, as was also true in an analysis of their *gag* genes.

Phylogenetic analysis of primate endogenous type D sequences. To gain insight into the origin of the retroviruses sequenced, tests were done to determine whether the 23.1 and 25.2 proviral clones represent exogenous or endogenous vi-

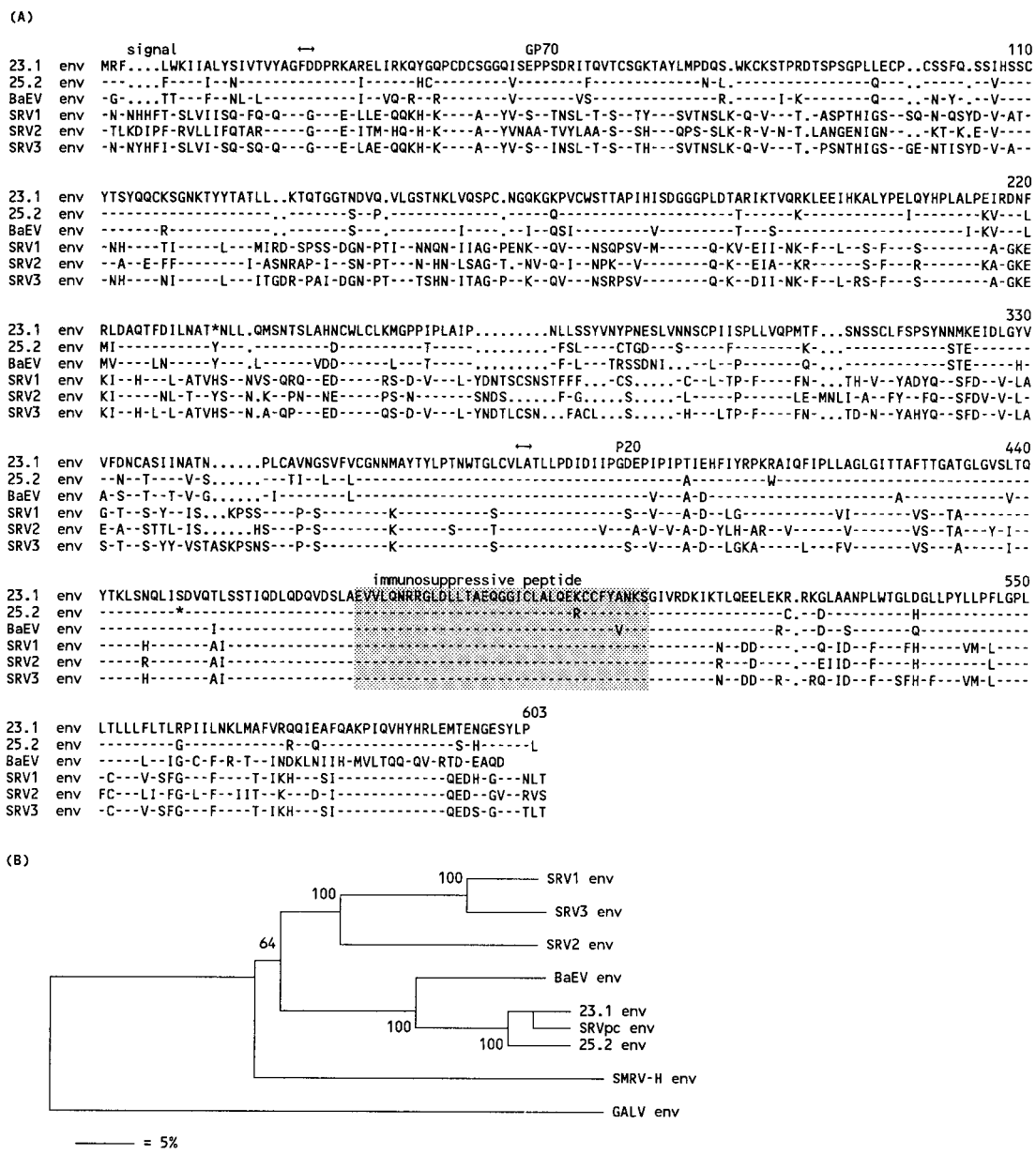

FIG. 5. (A) Alignment of *env* proteins derived from type D virus isolates 23.1 and 25.2, the *env* protein of BaEV, and the exogenous type D virus *env* proteins of SRV-1, SRV-2, and SRV-3. The signal peptide, GP70, and P20 proteins are indicated by arrows; the putative immunosuppressive peptide has been shaded; premature stop codons in clones 23.1 and 25.2 are indicated by an asterisk. (B) NJ tree based upon derived protein sequences for the *env* genes of clones 23.1 and 25.2 and primate type C and type D viruses. Bootstrap values for 100 replicated trees are indicated. SRV1 to SRV3 are exogenous Old World type D viruses, SMRV-H is a New World type D virus, GALV is an exogenous type C virus, and BaEV is a chimeric type C/type D endogenous virus. The placement of the baboon isolate SRVpc was taken from an NJ analysis of an approximately 580-nt fragment of Old World type D viruses and BaEV. Gaps introduced for optimal alignment are indicated by dots; identical amino acids are indicated by dashes.

ruses of baboons. As PCR results can be confusing due to the amplification of homologous genes belonging to other endogenous-virus families (46), primers should preferentially amplify an easily identifiable part of the virus. We decided to locate the upstream and downstream primers in different genes; e.g., the 5′ primer is located in the *pol* gene and the 3′ primer is in the *env* gene, and together they amplify 764 nt of viral sequence. In this way, PCR reactions are specific for 23.1 and 25.2 viral sequences and do not amplify, e.g., BaEV *env* genes. Animals belonging to four subspecies of baboons were all found to be PCR positive with this primer pair, and subsequent sequencing showed that fragments homologous to both the 23.1 and 25.2 viruses were present. Phylogenetic analysis of the obtained sequences indicated that 23.1- and 25.2-related sequences constitute two separate clusters (bootstrap value separating the clusters, 100), suggesting that they belong to different strains of the same virus (result not shown). It is less likely that one of the clusters represents integrations that have acquired multiple substitutions.

Additional PCR amplification of DNA samples from Old World primates, including chimpanzees and humans, showed that the 23.1 and 25.2 proviral genomes represent ancient

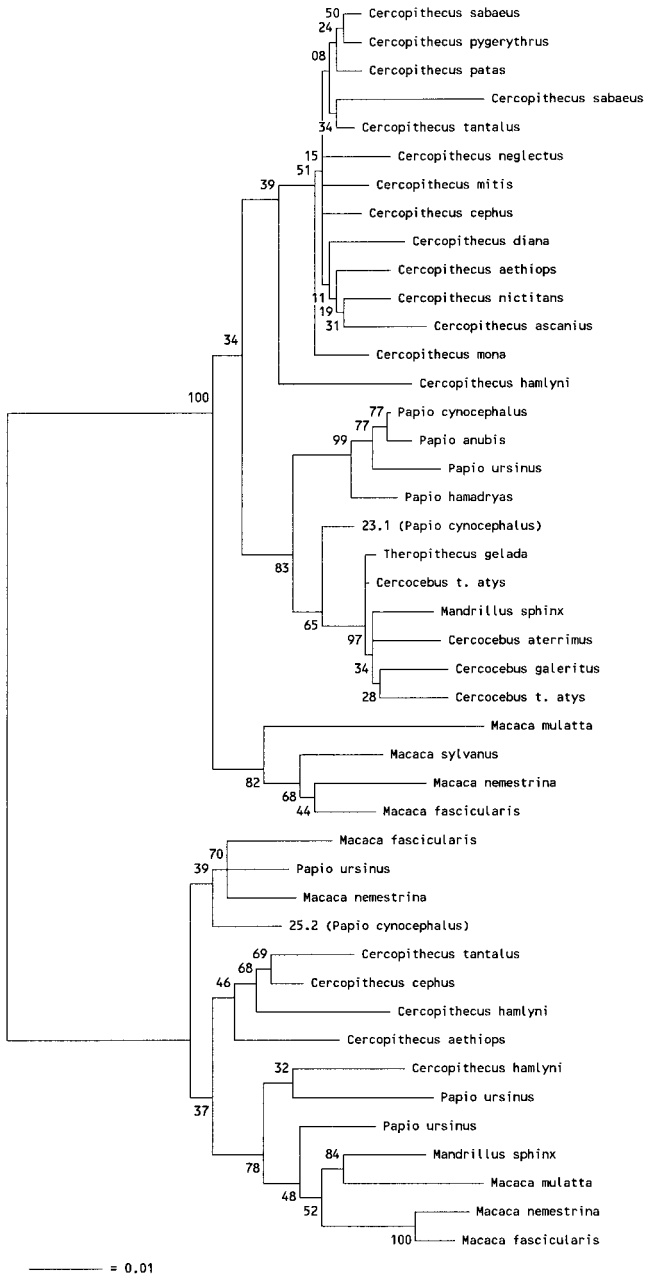

FIG. 6. NJ tree based upon approximately 764-nt *pol-env* fragments obtained from 23 species of African monkeys and three Asian macaque species. A single clone was analyzed per species, although more were sequenced. However, some divergent sequences belonging to the same species were also analyzed. The homologous sequences from baboon clones 23.1 and 25.2 were included in the analysis. Bootstrap values for 100 replicated trees are indicated.

integrations which are found in all species of Old World monkeys tested but not in humans and apes. Phylogenetic analysis (Fig. 6) suggests that the viral genomes have coevolved with their respective hosts, as especially for the 23.1 viruses there is an interesting similarity between the virus and host evolutionary trees. A host tree is presented by van der Kuyl et al. (46, 47). All clone 23.1-related fragments amplified from *Cercopithecus* monkeys form a cluster, although sequences derived from single species do not cluster together (result not shown). This could be due to the comparison of different copies. Also,

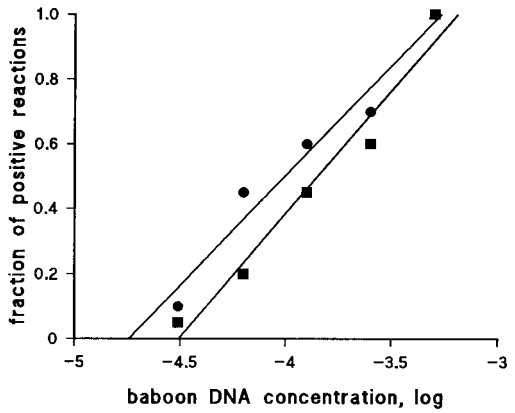

FIG. 7. Plot of the fraction of positive nested PCRs for the 23.1 and 25.2 proviruses as a function of the baboon genomic DNA concentration. The results for two male *Papio hamadryas* baboons are shown. Copy number calculations were based on duplicate experiments for each animal.

the time course might be too short to provide strong virus-host associations. After integration, the viral genome depends on the cellular machinery, which has a much lower mutation rate than the retrovirus itself (10), for replication. A second stable cluster (bootstrap value, 83) is formed by clone 23.1-related sequences amplified from monkeys belonging to the *Papionini* family (baboons, geladas, mangabeys, mandrills, and macaques). In the tree shown, the macaque viruses form a separate cluster but the bootstrap value is not significant, and in some analyses, macaque viruses were clustered with the other *Papionini* viruses. No clustering of virus sequences from a certain species was observed (results not shown), again suggesting that the time elapsed since the species evolved from their common ancestor was too short to establish strong virus-host associations.

From the macaque results it can be concluded that there is no specific virus variant associated with a particular geographic distribution, as the barbary macaque is African and other macaque species are from Asia but their endogenous type D viruses cluster together.

A clear division into *Cercopithecus* and *Papionini* viruses cannot be observed in the clone 25.2-related fragments (Fig. 6). Two small clusters, composed of sequences belonging to each of these families, are found, but the bootstrap values are low and a third cluster containing viruses from both groups is also present.

Copy numbers. Copy numbers of the 23.1 and 25.2 proviruses in the baboon genome were estimated by limited dilution and nested PCR. The PCR reactions were optimized so that the nested PCR was able to detect a single genome copy. Genomic DNA from two male *Papio hamadryas* baboons was used as input in the limiting-dilution procedure. For the calculations it was assumed that a baboon cell contains the same amount of genomic DNA as a human cell (6 pg of DNA/cell). The limiting-dilution method followed was as described by Ouspenskaia et al. (31). Duplicate experiments suggested that the number of SERV genomes in the male *Papio hamadryas* diploid genome is in the range of 142 to 235 provirus copies per baboon genome (Fig. 7).

DISCUSSION

The complete sequence of a novel type of endogenous retrovirus presented in this paper and its presence in multiple monkey species demonstrate that type D retroviruses are an-

cient and ubiquitous in the primate world. The endogenous 23.1-type D sequence is remarkably intact and shows interesting features when compared to known endogenous and exogenous viruses. At this moment there is no clear classification for endogenous viruses (9), so we propose that this new provirus be named simian endogenous retrovirus type D. The LTRs of the virus have no homology to known viral LTRs except for the 3′-most 140 nucleotides, which can be easily aligned with the 3′ ends of SRV LTRs. Also, they are much longer than SRV LTRs (484 nt versus 345 to 346 nt). The PBS, which is complementary to $tRNA_3^{Lys}$, is different from those of type D viruses. This tRNA is generally used by type B viruses and by the lentiviruses, while $tRNA_{1,2}^{Lys}$ is involved in the replication of type D viruses, including SMRV, and several others (for a review, see reference 24). The organization of the *gag*, protease, and *pol* genes of the 23.1 isolate is identical to that of these genes in type D viruses, and there is a high degree of homology between the sets of genes of these viruses at both the nucleotide and the amino acid levels. However, major differences are found in the *env* gene. The GP70 protein encoded by this gene has almost no similarity to type D virus GP70 but is homologous to the BaEV protein. In the P20 transmembrane protein, which is encoded by the C-terminal part of the *env* gene, there is again considerable homology with the type D virus proteins. However, BaEV P20 is also very similar to the type D P20 protein. Possibly the P20 protein is involved in receptor binding, as it has been shown that type D viruses, BaEV, and the cat virus RD114 (which has an *env* gene homologous to that of BaEV) have the same cellular receptor (37). The gene encoding this receptor is also present in the human genome (38).

Earlier we showed that BaEV genomes are found only in a distinct subset of African monkeys and that germ line integrations occurred approximately 24,000 to 400,000 years ago (46), which is quite recent in evolutionary terms. Suggestions that BaEV is a recombinant virus have been made before (20). BaEV probably has a genome which is a product of recombination between the *gag-pol* region of a type C virus and the *env* gene of the endogenous type D virus presented here. Because endogenous type D proviruses are found in all monkeys of the *Papionini* and *Cercopithecini* tribes, and phylogenetic analysis has shown that the virus sequences have coevolved with their hosts, it is most likely that SERV is ancestral to BaEV. In contrast, BaEV genomes are found in only a limited set of African monkeys (46). This finding indicates that in the recent past a new primate retrovirus evolved by recombination involving the *env* gene of an endogenous primate virus which was indigenous to the species later infected by the recombinant. The history of this new retrovirus infection can be deduced from the characteristics of its distribution in the form of endogenous proviruses in extant monkey species (46, 48). The other parent of BaEV could be a primate type C virus. Several type C viruses have been found in primates, including the Old World viruses GALV from gibbons (43) and MAC-1 and MMC-1 from macaques and the New World isolate OMC-1, for which related endogenous sequences have been observed in owl monkey genomic DNA (44). Nothing is known about the pathogenicity of either BaEV or SERV. The exogenous type D viruses found in captive macaques could be products of recombination between the SERV *gag-pol* genes and a GP70 *env* protein gene of unknown origin. SRVs are pathogenic in macaques, and multiple disease symptoms have been observed (28, 40). The immunodeficiency syndrome seen most often was attributed to a highly conserved putative immunosuppressive peptide encoded by the *env* gene. The hypothesis that the immune system of the macaque, in which BaEV genomes are not present, is susceptible to the immunosuppressive effect of this peptide is interesting in light of the new finding that macaques do harbor SERV genomes, in which the immunosuppressive peptide sequence is highly conserved. It is possible that this peptide is not involved in immunosuppression or that susceptibility to exogenous type D viruses and subsequent immunosuppression are correlated with individual expression levels of SERV proviral genomes. Alternatively, this domain could be involved in receptor binding, as sequences N and C terminal to the putative immunosuppressive peptide are completely conserved (Fig. 5A).

Earlier hybridization studies had already suggested that type D-related endogenous sequences could be found in several species of Old World monkeys (2, 4). PCR and sequence analysis of Old World monkey DNA showed that SERV is present in all *Papionini* and *Cercopithecini* (sub)species, suggesting that the virus integrated into the germ line of a common ancestor. As the provirus is not detectable in apes and humans, integration must have taken place after separation of the monkey branch from the hominoids, which occurred approximately 36 million years ago (MYA) (27). A more exact time point cannot be given until *Colobus* genomic DNA has been analyzed, as the *Colobinae* were established as distinct species by 9 MYA, suggesting that separation from the common ancestor of the *Cercopithecidae* occurred before that time point. In preliminary experiments using *Colobus guereza* genomic DNA, 23.1 and/or 25.2 sequences could not be amplified, suggesting that the integration of SERV strains in the primate germ line occurred later than 9 MYA.

A more distantly related type D virus, SMRV, has been isolated from New World squirrel monkeys (16). Only partial sequences are known for this virus (LTR and *gag* and *pol* genes), although a complete sequence is known for a human cell line contaminant of this monkey virus. SMRV placement in phylogenetic trees is consistent with a long-time separation from its counterparts, the Old World monkeys. Type D viruses have probably been associated with primates since they evolved. SMRV then represents the New World type D virus branch as the ancestor of its host separated from the Old World primates as early as 55 MYA (27), suggesting that type D viruses have entered the primate germ line at least twice.

SERV genomes sequenced are surprisingly intact after residing in the primate genome for such a long time. As there are approximately 142 to 235 integrations in the baboon genome, it is not unlikely that at least one of them contains ORFs for all proteins and can be expressed. Also, recombination between different proviral genomes could result in viable virus. There are several reports on the (accidental) recovery of type D virus particles from human cell lines (1, 22, 30). No endogenous human type D virus could be found in these cell lines, and although contamination with macaque type D viruses could be ruled out, it is possible that the cultures had been contaminated with (often-used) monkey cell lines (e.g., Vero, CV-1, or Cos) in which endogenous type D virus genomes are present. This would indicate that endogenous type D monkey viruses can be, and have been, expressed. SRV-3 can productively infect human cells with a low input of virus particles, (12), and the virus is also able to spread rapidly in animal populations (as high as 90% seropositivity in primate centers in a few years). BaEV particles can be easily obtained by cocultivating baboon tissue with human cells (3). The recent observation of infection of baboons in a primate center with a type D virus (14) is another finding suggesting that endogenous type D sequences can be expressed. Although the authors state that the infection is due to a recent cross-species transmission of SRV from macaques kept at the same center, phylogenetic analysis of the

obtained *gag* and *env* nucleotide fragments showed that the baboon isolate SRVpc is highly related to the SERV isolate 23.1 from baboons (Fig. 2B and 5B). An additional explanation could be that SRVpc is a recent product of recombination between an exogenous macaque virus and endogenous baboon sequences. Seropositivity for antibodies directed against type D virus proteins is regularly found in monkeys, both in the wild and in captivity (12, 14, 18, 26). Neutralizing antibodies were observed to persist throughout life, suggesting that type D viruses can establish persistent infections. Isolation of virus particles from captive macaques has been successful (26). The baboon isolate SRVpc can induce formation of multinucleated syncytia in human cells after cocultivation with baboon peripheral blood mononuclear cells. It is not known if the antibodies are directed against endogenous proteins or currently unrecognized SRV variants, but Western blot reactions of talapoin monkey sera deviated from those of control SRV (18), and the same was true for immunoblots from seropositive baboons when tested with SRV2 antiserum (14).

From these results, it can be concluded that different full-length endogenous proviruses exist in multiple copies in the baboon genome and most likely also in other monkey species. Phylogenetic analysis suggests that SERV is an ancient type D virus of Old World monkeys but can still be activated, as the baboon isolate SRVpc was shown to be closely related to the 23.1 isolate of SERV. In the recent past, multiple SERV sequence recombinants have evolved. The *env* gene of SERV has been incorporated into a type C virus to give rise to BaEV, which has been spreading among African monkeys since only 24,000 to 400,000 years ago. The recombinant SRV, containing SERV *gag* and *pol* genes and the *env* P20 gene, is presently active among Asian macaques, in which it is giving rise to an immunodeficiency syndrome (SAIDS).

ACKNOWLEDGMENT

This study was funded in part by the Institute of Virus Evolution and the Environment.

REFERENCES

1. **Asikainen, K., M. Vesanen, T. Kuittinen, and A. Vaheri.** 1993. Identification of human type D retrovirus as a contaminant in a neuroblastoma cell line. Arch. Virol. **129:**357–361.
2. **Barker, C. S., J. W. Wills, J. A. Bradac, and E. Hunter.** 1985. Molecular cloning of the Mason-Pfizer monkey virus genome: characterization and cloning of subgenomic fragments. Virology **142:**223–240.
3. **Benveniste, R. E., M. M. Lieber, D. M. Livingston, C. J. Sherr, and G. J. Todaro.** 1974. Infectious C-type virus isolated from a baboon placenta. Nature **248:**17–20.
4. **Benveniste, R. E., and G. J. Todaro.** 1977. Evolution of primate oncornaviruses: an endogenous virus from langurs (*Presbytis* spp.) with related virogene sequences in other Old World monkeys. Proc. Natl. Acad. Sci. USA **74:**4557–4571.
5. **Bohannon, R. C., L. A. Donehower, and R. J. Ford.** 1991. Isolation of a type D retrovirus from B-cell lymphomas of a patient with AIDS. J. Virol. **65:** 5663–5672.
6. **Boller, K., H. König, M. Sauter, N. Mueller-Lantzsch, R. Löwer, J. Löwer, and R. Kurth.** 1993. Evidence that HERV-K is the endogenous retrovirus sequence that codes for the human teratocarcinoma-derived retrovirus HTDV. Virology **196:**349–353.
7. **Boom, R., C. J. A. Sol, M. M. M. Salimans, C. L. Jansen, P. M. E. Wertheim-van Dillen, and J. van der Noordaa.** 1990. Rapid and simple method for purification of nucleic acids. J. Clin. Microbiol. **28:**495–503.
8. **Bray, M., S. Prasad, J. W. Dubay, E. Hunter, K.-T. Jeang, D. Rekosh, and M.-L. Hammarskjöld.** 1994. A small element from the Mason-Pfizer monkey virus genome makes human immunodeficiency virus type 1 expression and replication Rev-independent. Proc. Natl. Acad. Sci. USA **91:**1256–1260.
9. **Coffin, J. M.** 1992. Structure and classification of retroviruses, p. 19–49. *In* J. A. Levy (ed.), The Retroviridae, vol. I. Plenum Press, New York, N.Y.
10. **Domingo, E., and J. J. Holland.** 1988. High error rates, population equilibrium and evolution of RNA replication systems, p. 3–36. *In* E. Domingo, J. J. Holland, and P. Ahlquist (ed.), RNA genetics. CRC Press, Boca Raton, Fla.
11. **Elder, J. H., D. L. Lerner, C. S. Hasselkus-Light, D. J. Fontenot, E. Hunter, P. A. Luciw, R. C. Montelaro, and T. R. Phillips.** 1992. Distinct subsets of retroviruses encode dUTPase. J. Virol. **66:**1791–1794.
12. **Fine, D. L., G. C. Clarke, and L. O. Arthur.** 1979. Characterization of infection and replication of Mason-Pfizer monkey virus in human cell cultures. J. Gen. Virol. **44:**457–469.
13. **Gardner, M. B., M. Endres, and P. Barry.** 1994. The simian retroviruses SIV and SRV, p. 133–276. *In* J. A. Levy (ed.), The Retroviridae, part 3. Plenum Press, New York, N.Y.
14. **Grant, R. F., S. K. Windsor, C. K. Malinak, C. R. Bartz, A. Sabo, R. E. Benveniste, and C.-C. Tsai.** 1995. Characterization of infectious type D retrovirus from baboons. Virology **207:**292–296.
15. **Harrison, G. P., E. Hunter, and A. M. L. Lever.** 1995. Secondary structure model of the Mason-Pfizer monkey virus 5′ leader sequence: identification of a structural motif common to a variety of retroviruses. J. Virol. **69:**2175–2186.
16. **Heberling, R. L., S. T. Barker, S. S. Kalter, G. C. Smith, and R. J. Helmke.** 1977. Oncornavirus: isolation from a squirrel monkey (*Saimiri sciureus*) lung culture. Science **195:**289–292.
17. **Higgins, D. G., and P. M. Sharp.** 1988. CLUSTAL: a package for performing multiple sequence alignments on a microcomputer. Gene **73:**237–244.
18. **Ilyinskii, P., M. Daniel, N. Lerche, and R. Desrosiers.** 1991. Antibodies to type D retrovirus in talapoin monkeys. J. Gen. Virol. **72:**453–456.
19. **Jacks, T., and H. E. Varmus.** 1985. Expression of the Rous sarcoma virus *pol* gene by ribosomal frameshifting. Science **23:**1237–1242.
20. **Kato, S., K. Matsuo, N. Nishimura, N. Takahashi, and T. Takano.** 1987. The entire nucleotide sequence of baboon endogenous virus DNA: a chimeric genome structure of murine type C and simian type D retroviruses. Jpn. J. Genet. **62:**127–137.
21. **Kimura, M. A.** 1980. A simple method for estimating evolutionary rate of base substitution through comparative studies of nucleotide sequences. J. Mol. Evol. **16:**111–120.
22. **Krause, H., V. Wunderlich, and W. Uckert.** 1989. Molecular cloning of a type D retrovirus from human cells (PMFV) and its homology to simian acquired immunodeficiency type D retroviruses. Virology **173:**214–222.
23. **Kumar, S., K. Tamura, and M. Nei.** 1993. MEGA: Molecular Evolutionary Genetics Analyses, version 1.0. Pennsylvania State University, University Park, Pa.
24. **Leis, J., A. Aiyar, and D. Cobrinik.** 1993. Regulation of initiation of reverse transcription of retroviruses, p. 33–47. *In* A. M. Skalka and S. P. Goff (ed.), Reverse transcriptase. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
25. **Lerner, D. L., P. C. Wagaman, T. R. Phillips, O. Prospero-Garcia, S. J. Henriksen, H. S. Fox, F. E. Bloom, and J. H. Elder.** 1995. Increased mutation frequency of feline immunodeficiency virus lacking functional deoxyuridine-triphosphatase. Proc. Natl. Acad. Sci. USA **92:**7480–7484.
26. **Lowenstine, L. J., N. C. Pedersen, J. Higgins, K. C. Pallis, A. Uyeda, P. Marx, N. W. Lerche, R. J. Munn, and M. B. Gardner.** 1986. Seroepidemiologic survey of captive Old-World primates for antibodies to human and simian retroviruses, and isolation of a lentivirus from sooty mangabeys (*Cercocebus atys*). Int. J. Cancer **38:**563–574.
27. **Martin, R. D.** 1993. Primate origins: plugging the gaps. Nature **363:**223–234.
28. **Marx, P. A., D. H. Maul, K. G. Osborn, N. W. Lerche, P. Moody, L. J. Lowenstine, R. V. Henrickson, L. O. Arthur, R. V. Gilden, M. Gravell, W. T. London, J. L. Sever, J. A. Levy, R. J. Munn, and M. J. Gardner.** 1984. Simian AIDS: isolation of a type D retrovirus and transmission of the disease. Science **223:**1083–1086.
29. **McGeoch, D. J.** 1990. Protein sequence comparisons show that the 'pseudoproteases' encoded by poxviruses and certain retroviruses belong to the deoxyuridine triphosphatase family. Nucleic Acids Res. **18:**4105–4110.
30. **Oda, T., S. Ikeda, S. Watanabe, M. Hatsushika, K. Akiyama, and F. Mitsunobo.** 1988. Molecular cloning, complete nucleotide sequence, and gene structure of the provirus genome of a retrovirus produced in a human lymphoblastoid cell line. Virology **167:**468–476.
31. **Ouspenskaia, M. V., D. A. Johnston, W. M. Roberts, Z. Estrov, and T. F. Zipf.** 1995. Accurate quantification of residual B-precursor acute lymphoblastic leukemia by limiting dilution and a PCR based detection system: a description of the method and the principles involved. Leukemia **9:**321–328.
32. **Poch, O., I. Sauvaget, M. Delarue, and N. Tordo.** 1989. Identification of four conserved motifs among the RNA-dependent RNA polymerase encoding elements. EMBO J. **8:**3867–3874.
33. **Power, M. D., P. A. Marx, M. L. Bryant, M. B. Gardner, P. J. Barr, and P. A. Luciw.** 1986. Nucleotide sequence of SRV-1, a type D simian acquired immune deficiency syndrome retrovirus. Science **231:**1567–1572.
34. **Rabin, H., C. V. Benton, M. A. Tainsky, N. R. Rice, and R. V. Gilden.** 1979. Isolation and characterization of an endogenous type C virus of rhesus monkeys. Science **204:**841–842.
35. **Rizvi, T. A., K. A. Lew, E. C. Murphy, Jr., and R. D. Schmidt.** 1996. Role of Mason-Pfizer monkey virus (MPMV) constitutive transport element (CTE) in the propagation of MPMV vectors by genetic complementation using homologous/heterologous *env* genes. Virology **224:**517–532.
36. **Sherwin, S. A., and G. J. Todaro.** 1979. A new endogenous primate type C

virus isolated from the Old World monkey *Colobus polykomos*. Proc. Natl. Acad. Sci. USA **76:**5041–5045.
37. **Sommerfelt, M. A., and R. A. Weiss.** 1990. Receptor interference groups of 20 retroviruses plating on human cells. Virology **176:**58–69.
38. **Sommerfelt, M. A., B. P. Williams, A. McKnight, P. N. Goodfellow, and R. A. Weiss.** 1990. Localization of the receptor gene for type D simian retroviruses on human chromosome 19. J. Virol. **64:**6214–6220.
39. **Sonigo, P., C. Barker, E. Hunter, and S. Wain-Hobson.** 1986. Nucleotide sequence of Mason-Pfizer monkey virus: an immunosuppressive D-type retrovirus. Cell **45:**375–385.
40. **Stromberg, K., R. E. Benveniste, L. O. Arthur, H. Rabin, W. E. Giddens, Jr., H. D. Ochs, W. R. Morton, and C-C. Tsai.** 1984. Characterization of exogenous type D retrovirus from a fibroma of a macaque with simian AIDS and fibromatosis. Science **224:**289–292.
41. **Thayer, R. M., M. D. Power, M. L. Bryant, M. B. Gardner, P. J. Barr, and P. A. Luciw.** 1987. Sequence relationships of type D retroviruses which cause simian acquired immunodeficiency syndrome. Virology **157:**317–329.
42. **Todaro, G. J., R. E. Benveniste, S. A. Sherwin, and C. J. Sherr.** 1978. MAC-1, a new genetically transmitted type C virus of primates: "low frequency" activation from stumptail monkey cell cultures. Cell **13:**775–782.
43. **Todaro, G. J., M. M. Lieber, R. E. Benveniste, C. J. Sherr, C. J. Gibbs, and D. C. Gajdusek.** 1975. Infectious primate type C viruses: three isolates belonging to a new subgroup from the brains of normal gibbons. Virology **67:**335–343.
44. **Todaro, G. J., C. J. Sherr, A. Sen, N. King, M. D. Daniel, and B. Fleckenstein.** 1978. Endogenous New World primate type C viruses isolated from owl monkey (*Aotus trivirgates*) kidney cell line. Proc. Natl. Acad. Sci. USA **75:**1004–1008.
45. **van der Kuyl, A. C., J. T. Dekker, and J. Goudsmit.** 1995. Full-length proviruses of baboon endogenous virus (BaEV) and dispersed BaEV reverse transcriptase retroelements in the genome of baboon species. J. Virol. **69:** 5917–5924.
46. **van der Kuyl, A. C., J. T. Dekker, and J. Goudsmit.** 1995. Distribution of baboon endogenous virus among species of African monkeys suggests multiple ancient cross-species transmissions in shared habitats. J. Virol. **69:**7877–7887.
47. **van der Kuyl, A. C., C. L. Kuiken, J. T. Dekker, and J. Goudsmit.** 1995. Phylogeny of African monkeys based upon the mitochondrial 12S rRNA gene. J. Mol. Evol. **40:**173–180.
48. **van der Kuyl, A. C., J. T. Dekker, and J. Goudsmit.** 1996. Baboon endogenous virus evolution and ecology. Trends Microbiol. **4:**455–459.